

What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers

Year: 2021 | Citations: 128 | Authors: Boseop Kim, Hyoungseok Kim, Sang-Woo Lee, Gichang Lee, Donghyun Kwak

Abstract

GPT-3 shows remarkable in-context learning ability of large-scale language models (LMs) trained on hundreds of billion scale data. Here we address some remaining issues less reported by the GPT-3 paper, such as a non-English LM, the performances of different sized models, and the effect of recently introduced prompt optimization on in-context learning. To achieve this, we introduce HyperCLOVA, a Korean variant of 82B GPT-3 trained on a Korean-centric corpus of 560B tokens. Enhanced by our Korean-specific tokenization, HyperCLOVA with our training configuration shows state-of-the-art in-context zero-shot and few-shot learning performances on various downstream tasks in Korean. Also, we show the performance benefits of prompt-based learning and demonstrate how it can be integrated into the prompt engineering pipeline. Then we discuss the possibility of materializing the No Code AI paradigm by providing AI prototyping capabilities to non-experts of ML by introducing HyperCLOVA studio, an interactive prompt engineering interface. Lastly, we demonstrate the potential of our methods with three successful in-house applications.